

# External Validity, Why Art Thou Externally Valid? Recent Studies of Attraction Provide Three Theoretical Answers

Paul W. Eastwick\*, Lucy L. Hunt and Lisa A. Neff

University of Texas at Austin

---

## Abstract

Some studies have better external validity than others, but why? Recent studies in the domain of interpersonal attraction have been tackling this question by documenting how people respond differently to hypothetical versus live interactions. In live interactions, people tend to report their experienced emotions, they evaluate others using a low-level concrete construal, and they attempt to implement the goal of having a pleasant interaction. In hypothetical scenarios, people forecast their emotions, they evaluate others using a high-level abstract construal, and they deliberate about others' positive and negative features. By situating the hypothetical versus live interaction distinction within the framework of strong preexisting theories (i.e., affective forecasting, construal-level theory, mindset theory), this research reinforces the idea that there is nothing inherently invalid about laboratory studies that are cosmetically dissimilar from real life. Nevertheless, it remains highly problematic to generalize findings to a setting that elicits a countervailing set of psychological processes.

---

Every so often, the field of psychology stands on a moonlit balcony and reflects on whether our research reveals something important about how people experience real life. Frequently, this soul-searching centers on the concept of external validity, which refers to the extent to which the results of an experiment generalize to samples and settings beyond those examined in the original study (Campbell, 1957). Over the decades, scholars have debated the severity of the field's external validity troubles, with some suggesting the lack of external validity to be a serious shortcoming of our work (Baumeister, Vohs, & Funder, 2007; Carlson, 1984; Cialdini, 2009; Mitchell, 2012) and others more sanguine about the real-world importance of the phenomena that psychologists examine (Anderson, Lindsay, & Bushman, 1999; Kenrick, 1986; Mook, 1983).

Currently, the consensus view about external validity seems to be as follows. Firstly, the observations of Mook (1983) remain relevant and highly influential: The critical external validity question is whether the psychological *process* of interest generalizes from the laboratory to the real world (i.e., *psychological realism*; Aronson, Wilson, & Akert, 1994; Brewer, 2000). Therefore, scholars will often have good reason to perform experiments that are cosmetically dissimilar from real life settings, especially if the experiment tests a critical element of a psychological theory or process. Secondly, a clever meta-analytical approach pioneered by Anderson et al. (1999) has documented a substantial association between the laboratory effect size and the field effect size of the same psychological process across a wide array of social psychological topics (Anderson et al., 1999,  $r = .73$ ; Mitchell, 2012,  $r = .53$ ). In other words, laboratory experiments tend to produce effects that have a similar (relative) size to the effects that are documented in field settings. Indeed, the traditional tradeoff between the experimental control of the laboratory and the realism of the field is far from absolute, as many studies have successfully blended the two approaches (Reis & Gosling, 2010). Thirdly, room for improvement remains. Despite these substantial correlations across all topics, laboratory

and field effect sizes do align weakly in certain domains (Mitchell, 2012); for example, sex differences in the use of an interpersonally oriented leadership style are much larger in laboratory settings (i.e., women are more interpersonally oriented than men) than in the actual workplace (Eagly & Johnson, 1990). Furthermore, many have argued that social psychological scholarship will achieve greater influence if researchers conduct studies that have a clear and immediate impact on people's lives (Cialdini, 2009) and assess actual behavior in lieu of self-report responses to hypothetical scenarios (Baumeister et al., 2007).

Assuming that scholars wish to take up these challenges, what is the best way to conduct studies that have real-world relevance? Why might some studies prove to have strong external validity, whereas other findings end up largely confined to artificial laboratory settings? On this point, the existing literature offers little guidance. One possibility is that the college student samples frequently used in laboratory research may not generalize to the population at large, an issue that has received extensive treatment elsewhere (Henrich, Heine, & Norenzayan, 2010; Henry, 2008, Sears, 1986). Other suggestions for improving external validity have focused on the development of paradigms that “feel real” to the participants or sample from real-world stimuli (Aronson et al., 1994; Dhami, Hertwig, & Hoffrage, 2004; Mitchell, 2012). Yet, these perspectives do not satisfactorily address *why* some psychological processes are more real than others, other than in a tautological sense (i.e., the circular notion that a paradigm elicits real psychological processes if the paradigm comes from the real world). In essence, despite seasonal hand-wringing about external validity, the literature directly addressing this topic offers no theoretical justification for predicting *a priori* that a study should or should not be externally valid.

In this article, we propose a novel way to address this question, which is that scholars could draw from existing psychological theories to predict differences between a laboratory version of a paradigm and an externally valid, field-like analog. In fact, over the past five years, several independent lines of research on interpersonal attraction have used exactly this strategy. These studies have focused specifically on the distinction between evaluating a hypothetical person whom one has never met (e.g., imaginary scenarios or lists of traits), which rarely occurs outside the laboratory, and evaluating a live person (e.g., face-to-face interaction), which more closely mimics the way that people meet each other in the real world (Funder, 1995).<sup>1</sup> Although many of these studies did not explicitly intend to address the issue of external validity, we suggest that this new attraction research is among the first to illuminate the concept of external validity in a theoretically generative manner. Ironically, the picture that emerges reinforces the suggestion that external invalidity is not the true enemy of good research (Mook, 1983); rather, researchers err by generalizing their results to a situation that is likely to elicit a countervailing psychological process.

The next section of this article reviews three theoretical perspectives that have demonstrated great promise in helping researchers to understand why a given study will or will not exhibit external validity. These three perspectives are affective forecasting/empathy gap theories (Loewenstein, 1996, 2005; Wilson & Gilbert, 2003), construal level theory (Liberman, Trope, & Stephan, 2007; Trope & Liberman, 2003, 2010), and mindset theory (Gollwitzer, 1990, 2012; Gollwitzer & Bayer, 1999). These three theories delineate three different (but related) dimensions of psychological functioning<sup>2</sup>, and they offer justification for the hypothesis that one end of the dimension may be more likely than the other to characterize field-based, “real-world” research. The subsequent section reviews several independent lines of attraction research – two from the intergroup domain and three investigating classic attraction principles – that have drawn from one or more of these theoretical perspectives to document different psychological processes taking place in hypothetical versus live interactions. This work finds that hypothetical interactions are often characterized by forecasted

emotions, abstract construals, and deliberative mindsets, whereas live interactions are often characterized by experienced emotions, concrete construals, and implemental mindsets. The final section of the article stresses that both ends of all three dimensions are psychologically valid and real; in fact, forecasting, abstract thinking, and deliberation may be uniquely human mental features that endowed our ancestors with an impressive evolutionary advantage. Nevertheless, researchers should endeavor to understand and predict how the very nature of a research paradigm may cause participants' minds to inhabit a particular space along these dimensions.

## Theoretical Perspectives Relevant to the Hypothetical Versus Live Interaction Distinction

### *Affective forecasting*

The psychological processes involved in generating emotional judgments about the past or future can differ from the processes that affect actual emotional experiences in the present. After all, emotions are momentary: People experience their current feelings, but they cannot directly summon and experience a perfect facsimile of an emotion that they felt in the past or an emotion that they will feel in a future hypothetical situation (Robinson & Clore, 2002). Rather, people draw upon their emotional schemas – their beliefs about the kind and intensity of emotion that they will experience under particular circumstances – when completing emotional reports about the past or the future. A large corpus of work has documented differences between anticipated future emotions and experienced emotions, otherwise known as affective forecasting errors (for a review, see Wilson & Gilbert, 2003).

Affective forecasting errors can arise from a variety of sources. People's construal of an emotion-inducing event can be flawed such that people make forecasts about an event that is fundamentally different than the one that they actually experience; for example, women may overestimate the likelihood that they would reprimand a sexual harasser because they imagine a situation in which it will be easy to confront him (Woodzicka & LaFrance, 2001). In addition, people may possess incorrect theories about the emotions that the actual event will cause (McFarland, Ross, & DeCourville, 1989; Wilson, Laser, & Stone, 1982), and they exhibit *cold-to-hot empathy gaps* such that their forecasts in "cold" emotional states do not account for the "hot" motivational influences (e.g., pain, fear, or arousal) that affect their actual emotional experiences and behavior (Loewenstein, 2005). Finally, people also fail to realize that they possess a psychological immune system that responds to emotionally distressing events by helping them to reinterpret information in a way that mitigates their experienced negative emotions (Gilbert, Pinel, Wilson, Blumberg, & Wheatley, 1998).

Thus, the emotional self-reports that people generate when forecasting how they will feel may differ from the emotional reports that later characterize people's actual experiences. The idea that emotional self-reports can range from forecasted to experienced – and that these two kinds of reports are affected by different psychological processes – could prove relevant to the hypothetical versus live interaction distinction. Specifically, it is possible that people largely rely on affective forecasts when they make judgments about how they will react in hypothetical interactions with imaginary individuals, but they report their experienced emotions in the moment when reporting on a live, face-to-face interaction partner. For this reason, it is possible that people's emotional reactions will differ in response to hypothetical versus live interactions; later, we review two studies that drew from the affective forecasting perspective to document support for this hypothesis (Kawakami, Dunn, Karmali, & Dovidio, 2009; Mallett, Wilson, & Gilbert, 2008).

*Construal level theory*

A second theoretical perspective relevant to the distinction between evaluating a hypothetical versus a live interaction partner is construal level theory (Liberman et al., 2007; Trope & Liberman, 2003, 2010). According to this perspective, people can represent objects at either a high-level or a low-level construal. High-level construals are abstract, structured, and emphasize the core, essential features of someone or something; when people represent an object in a high-level construal, they extract the central gist of the object while ignoring peripheral details. In contrast, low-level construals are concrete, detailed, and emphasize contextualized, subordinate features; when people represent an object in a low-level construal, they focus on details and incorporate peripheral features. As an illustration, the same activity “taking an exam” can be represented at either a high-level (e.g., succeeding academically) or a low-level (e.g., writing with a pen) construal.

A core element of construal level theory is that people are more likely to use high-level, abstract construals when they are considering psychologically distant objects, and they are more likely to use low-level, concrete construals when they are considering psychologically near objects (Trope & Liberman, 2010). This distance dimension can literally refer to spatial distance (e.g., near vs. far), but it can also refer to distance in time (e.g., imminent vs. the past or future) as well as hypotheticality (e.g., real vs. imagined). This latter point is especially relevant to the present discussion: When people are considering hypothetical situations, they are likely to make use of high-level, abstract construals, and when they are considering information in the here-and-now, they are likely to make use of low-level, concrete construals.

This concrete versus abstract distinction has implications for the way that people evaluate others (Ledgerwood, Trope, & Liberman, 2010). When participants are using an abstract construal, they tend to make judgments that incorporate information from relevant schemas and ideologies, but when they are using a concrete construal, they rely to a greater extent on the immediate social context (Ledgerwood, Trope, & Chaiken, 2010). For example, people are more likely to believe that a person’s traits would direct his/her behavior across situations (Nussbaum, Trope, & Liberman, 2003) and more likely to infer that a person’s behavior connotes that he/she possesses a relevant trait (Rim, Uleman, & Trope, 2009) when in an abstract than a concrete mindset. This research suggests that people might be more likely to use traits – which are essentially abstract interpersonal schemas – as a basis for evaluating another person’s likability when using a high-level, abstract construal. Alternatively, people might use immediately accessible experiential information (e.g., momentary affect, gut-level evaluations, chemistry, or rapport; Frost, Chance, Norton, & Ariely, 2008) to evaluate another person when using a low-level, concrete construal. Two sets of studies that examined these possibilities are reviewed later in this article (Eastwick, Finkel, & Eagly, 2011; Park, Young, & Eastwick, 2013).

*Mindset theory*

Mindset theory (Gollwitzer, 1990, 2012; Gollwitzer & Bayer, 1999) is a third theoretical perspective that addresses the distinction between evaluating a hypothetical interaction partner and a live interaction partner. This theory proposes that different psychological mechanisms are relevant to the process of evaluating possible goals versus executing and implementing a particular selected goal. When people are evaluating several different goals and deciding which goal to pursue (i.e., the predecisional phase), they are in a *deliberative mindset*. People in a deliberative mindset attempt to accurately assess (a) the pros and cons of competing goals and (b) whether or not they are likely to achieve each goal successfully.

Alternatively, when people are pursuing a chosen goal (i.e., the postdecisional phase), they are in an *implemental mindset*. People in an implemental mindset focus on information that is useful in formulating plans that will lead to the successful completion of the selected goal, and they avoid information about whether the goal is desirable or attainable. In fact, participants often exhibit excessive optimism about the likelihood of achieving a goal, and this optimism frequently motivates continued goal pursuit in the face of setbacks (Taylor & Brown, 1988).

Mindsets have inertia; that is, when people are in a particular mindset, they carry that same processing style into subsequent yet unrelated tasks (Gollwitzer, 2012). Several studies in the romantic domain drew from this logic to illuminate the positive biases that people hold about their relationships (Gagné & Lydon, 2001, 2004; Gagné, Lydon, & Bartz, 2003). On average, people who are involved in a romantic relationship are likely to overestimate how long that relationship will ultimately last (MacDonald & Ross, 1999). This positive bias could reflect the fact that people typically approach their romantic relationships with an implemental mindset. Indeed, participants who were experimentally assigned to experience a deliberative mindset subsequently evaluated their relationships with an uncommon measure of objectivity (Gagné & Lydon, 2001; Gagné et al., 2003). Specifically, relative to participants in control or implemental conditions, participants in a deliberative mindset were (a) more likely to forecast accurately whether their relationships would break up over a 6-month period (Gagné & Lydon, 2001) and (b) more likely to generate commitment reports that accurately predicted later break-up status (Gagné et al., 2003). These findings suggest that the average person treats relationship maintenance as a goal that is currently under pursuit, but when in a deliberative mindset, people will weigh the pros and cons of their relationship.

But do implemental and deliberative mindsets also affect initial interactions? Just as people tend to approach their relationships with an implemental mindset – doggedly pursuing their chosen goal of maintaining positive feelings about their partners (Murray, Holmes, & Griffin, 1996) – perhaps people approach initial encounters (on average) with the goal of maintaining a smooth and pleasant interaction. However, hypothetical contexts may inspire a deliberative mindset that causes participants to weigh another individual's positive and negative traits. Below, we review a line of research that tested these hypotheses (Maniaci, Reis, Caprariello, Eastwick, & Finkel, 2012; Reis, Maniaci, Caprariello, Eastwick, & Finkel, 2011).

## Recent Attraction Research Examining the Hypothetical Versus Live Interaction Distinction

### *Intergroup interactions*

Humans are equipped with psychological barriers to intergroup contact, and these barriers often take the form of affective forecasting errors. For example, people may harbor inaccurately negative expectations about intergroup interactions because they tend to focus on dissimilarities with out-group members when considering such interactions in the abstract. However, Mallett et al. (2008) predicted that in an actual intergroup interaction, individuals would encounter similarities and therefore have a much more positive interaction than anticipated. Across four studies, the researchers assessed individuals' forecasted (i.e., hypothetical) and experienced (i.e., live) affective responses to intergroup interactions using daily diary records and interactions between participants and confederates. As predicted, participants forecasted that their experiences in intergroup interactions would be more negative than what they experienced when the interactions actually took place; in fact, participants tended to rate their actual intergroup experience just as positively as an average *intragroup* interaction. Furthermore, this affective forecasting error seemed to emerge because people



mistakenly focused on dissimilarities when considering hypothetical intergroup interactions: A manipulation that encouraged participants to focus on similarities when generating their forecasts successfully reduced the magnitude of the intergroup forecasting error. In other words, focusing on similarities rather than dissimilarities causes individuals' affective forecasts about intergroup interactions to become more positive and accurate. Taken together, the Mallett et al. (2008) findings demonstrate how excessively negative affective theories can cause judgments about hypothetical interactions to differ from actual affective experiences.

Other research has documented similarly faulty affective theories with respect to social deterrents to racism (Kawakami et al., 2009). In the abstract, majority group members are aware of the negative stigma associated with being a racist, yet some of them continue to exhibit racist behaviors. These researchers hypothesized that the tendency to punish ingroup members for failing to live up to abstract egalitarian norms is weaker than people anticipate, and so they conducted studies investigating individuals' affective reactions to a fellow ingroup member's racist acts. Drawing from affective forecasting perspectives as well as mindset theory, the researchers predicted that participants would consult their egalitarian beliefs when offered the opportunity to provide deliberative responses about encountering a hypothetical racist, but their nonconscious negative feelings about out-group members would drive their responses when confronted with a real-life racist. In two studies, White participants either (a) imagined interacting with or (b) actually interacted with a White confederate who either did or did not make a racist comment about a Black confederate. Whereas participants predicted that they would experience considerable distress in a hypothetical situation where the White individual made a racist comment, participants in the real-life interaction condition reported feeling little distress regardless of whether or not the White confederate made the racist comment. Moreover, participants in the hypothetical condition believed that they would be repulsed by the racist confederate and punish him by choosing the Black confederate as a task partner, but participants in the live interaction condition tended to choose the racist White confederate as a partner. This work suggests that, despite people's abstract endorsement of egalitarian ideals, people fail to punish racist individuals because their predictions about their own negative emotional experiences are far more intense than what they experience when the situation actually occurs.

### *Initial attraction principles*

Other recent research has found that people use different information to evaluate partners depending on whether they are imagining a hypothetical individual or evaluating live individuals after a face-to-face interaction. One such line of research examined the conditions under which people would romantically desire potential partners who match rather than mismatch their stated ideal partner preferences (Eastwick, Finkel, et al., 2011). Ideal partner preferences for traits are abstract constructs (Eastwick, Luchies, Finkel, & Hunt, in press), and according to construal level theory, participants' ideal partner preferences should predict their relational evaluations (e.g., romantic desire) to the extent that the evaluative context is abstract. However, in the concrete context of a live interaction, participants' romantic evaluations may not be associated with the extent to which the potential romantic partner matches their ideal partner preferences. In two studies, participants evaluated opposite-sex targets in an abstract context (i.e., a profile that listed the target's traits) and in a concrete context (i.e., a live, face-to-face interaction). As predicted, when participants evaluated potential romantic partners in the abstract, participants desired the partners who matched rather than mismatched their ideals. In contrast, when participants evaluated potential romantic partners in a concrete, live initial interaction, the match between the partner's traits

and the participants' ideal partner preferences no longer predicted romantic desire. In fact, evidence suggested that participants redefined the *meaning* of the potential partner's traits after the live interaction; for example, a participant who liked the live partner might have reinterpreted the partner's trait "outspoken" to mean "frank", whereas a participant who disliked the partner reinterpreted it to mean "tactless". That is, participants used the low-level cues present in the live interaction (e.g., their own gut-level evaluations of the partner) to contextualize and reinterpret their prior knowledge about the partner's abstract qualities. In short, people seem to use low-level experiential information (e.g., momentary affect; see also Eastwick, Eagly, Finkel, & Johnson, 2011) when evaluating romantic partners in a live, face-to-face setting, but they use high-level information when evaluating hypothetical romantic partners in settings lacking behavioral cues.

Another line of work found evidence for a similar disconnect between (a) participants' desire for certain qualities in a partner in the abstract and (b) participants' feelings in the moment about an actual partner who exhibits those qualities (Park et al., 2013). This series of studies examined men's reactions to intelligent women: Specifically, under what conditions are men attracted to women who outperform (vs. underperform) them on intelligence tests? Again, drawing from construal level theory, the authors hypothesized that in a psychologically distant, hypothetical scenario, men would consider their abstract beliefs that intelligent women are desirable. In contrast, in the psychologically close, low-level context of a live interaction, they hypothesized that the men's impressions would be dominated by the negative affective reaction that occurs as a result of being outperformed (Pleban & Tesser, 1981; Tesser, 1988). As predicted, when evaluating female targets in a hypothetical scenario or in another room (i.e., psychologically distant contexts), male participants reported greater romantic interest in women who outperformed them than underperformed them on an intelligence test. However, when evaluating female targets in live, face-to-face interactions (i.e., psychologically close contexts), male participants reported less romantic interest in women who outperformed than underperformed them. As with the studies cited above, these results suggest that as psychological distance decreases, individuals rely on proximal, situational cues (e.g., feelings of inferiority at being outperformed) in lieu of a partner's abstract, desirable characteristics.

A final relevant program of research examined the conditions under which familiarity leads to more or less liking for new acquaintances (Reis et al., 2011). Although classic studies have suggested that increased familiarity with another person leads to more liking for him/her (Festinger, Schachter, & Back, 1950; Moreland & Beach, 1992), other evidence has suggested the opposite effect: that familiarity breeds contempt. Consistent with the latter possibility, a recent set of studies by Norton and colleagues found that as the number of traits used to describe a hypothetical person increased, participants' liking for that person decreased (Norton, Frost, & Ariely, 2007). In an attempt to reconcile the Norton studies with the classic familiarity research, Reis et al. (2011) drew from mindset theory to suggest that participants might approach such a hypothetical target very differently than they approach live face-to-face interactions. Specifically, they suggested that evaluating lists of traits (as in Norton et al., 2007) activates a deliberative mindset and encourages participants to weigh the advantages and disadvantages of each trait, whereas live interactions activate an implemental mindset that motivates participants to work toward the goal of having a smooth, pleasurable interaction. In fact, Reis et al. (2011) found support for the idea that, in a live interaction paradigm, familiarity leads participants to report more liking for new acquaintances just as in the classic familiarity studies. A follow-up live interaction study (Maniaci et al., 2012) replicated this positive effect of familiarity on liking among participants primed with an implemental mindset; however, participants primed with a deliberative mindset tended to show

the negative effect of familiarity on liking found in the Norton et al. (2007) studies. This work suggests that people may approach lists of traits by weighing the pros and cons of those traits, whereas they approach live interactions with the intent to implement a smooth and pleasant interaction.

## Discussion

### *Implications*

The preceding review suggests that live interactions differ from hypothetical ones in several ways. In live interactions, people tend to report their actual emotional experiences; incorporate concrete, contextual details into their responses; and attempt to implement the goal of making the interaction proceed smoothly. In hypothetical scenarios, people tend to use schemas to forecast their emotions, consider abstract information such as ideals and ideologies, and deliberate about other people's positive and negative features. Three well-established theoretical perspectives – affective forecasting, construal level theory, and mindset theory – clearly predict the emergence of these psychological process differences between live and hypothetical paradigms (see Table 1).

Upon initial inspection, the live interaction paradigm would seem to have better external validity than the hypothetical scenario paradigm (Funder, 1995). People nearly always meet potential romantic partners face-to-face before they decide to form a romantic relationship (Finkel, Eastwick, Karney, Reis, & Sprecher, 2012)<sup>3</sup>; moreover, the complexities of real-life intergroup relations are inherently dyadic and were imperfectly assessed by past intrapersonal approaches (Shelton, 2000). In fact, several of the studies reviewed above used a live interaction paradigm because the researchers explicitly questioned whether people “in real-world social interaction” actually exhibit the same evaluations that had previously been demonstrated in hypothetical contexts (Eastwick, Finkel, et al., 2011; Reis et al., 2011). Thus, one lesson that emerges from these recent studies of attraction is that a prototypical externally valid study is likely to involve participants who are experiencing emotions, using concrete construals, or implementing goals. This work addresses the appeal of Mook (1983) to look past a study's cosmetic similarity to real life and focus instead on the generalizability of the relevant psychological processes to real life (i.e., psychological realism; Aronson et al., 1994; Brewer, 2000). In fact, experienced emotions, concrete construals, and implemental mindsets could comprise a substantial component of what psychological realism *means*, at least within the attraction realm.

However, there is a deeper conclusion lurking in these literatures, a conclusion that cautions against the knee-jerk dismissal of any mental process as psychologically unreal and externally invalid. The present work suggests that the hypothetical versus live interaction distinction covaries with certain psychological features; yet, even if people are more likely

**Table 1.** Theoretical distinctions between live and hypothetical interactions.

Theoretical dimension	Live interactions (field-based)	Hypothetical scenarios (laboratory-based)
Affective forecasting	Experienced	Forecasted
Construal-level theory	Concrete (low-level)	Abstract (high-level)
Mindset theory	Implemental	Deliberative



to use a low-level construal, experience emotions, and implement goals in live interactions than when considering hypothetical scenarios, it does not follow that abstraction, forecasting, and deliberation are invalid mental processes. After all, voting is an activity in which people transcend the here-and-now and consult their abstract beliefs and ideologies (Ledgerwood, Trope, & Liberman, 2010). Even though people only perform this behavior yearly (if that), it has profound real-world implications and psychologists must endeavor to understand it. In fact, forecasting and abstract thinking probably differentiate humans from the rest of the animal kingdom: These skills are likely to be new evolutionary innovations, emerging in the human lineage as recently as 50,000 years ago (Eastwick, 2009; Mithen, 1996). These newly evolved mental abilities provided humans with a degree of flexibility and foresight unprecedented in the animal kingdom, and these skills were likely critical in enabling humans to become the most successful hominids on the planet. Forecasting, abstraction, and deliberation are very real, and however frequently they emerge in daily life, their importance cannot be overstated.

Thus, whether we conceptualize external validity as cosmetic similarity to real life or as psychological realism, there is nothing inherently problematic about conducting an “externally invalid” study. A dilemma arises only when researchers study a set of mental processes in one setting and attempt to generalize those findings to a setting that elicits the countervailing set of processes. For example, abstract concepts should correlate with related abstract concepts: The abstract qualities that people rate highly in an ideal romantic partner are associated with abstract beliefs about the qualities of the self (Campbell, Simpson, Kashy, & Fletcher, 2001) and abstract anticipated future roles (Eagly, Eastwick, & Johannesen-Schmidt, 2009). However, abstract concepts may not translate to concrete contexts: As reviewed above, ideal partner preferences do not affect the way that people evaluate live opposite-sex individuals as they consider forming an actual romantic relationship (see also Eastwick & Finkel, 2008; Eastwick et al., in press). Ideal partner preferences are externally valid constructs, but they are abstract, and romantic relationship initiation is a low-level, affectively laden process (Eastwick, Eagly, et al., 2011; Simpson, Campbell, & Berscheid, 1986).

According to this line of reasoning, even the classic enemy of external validity – the demand characteristic – takes on a less insidious character. When demand characteristics are present, people might exhibit a particular evaluation or behavior simply because an experimenter wanted or expected it. Although this situation seems to exemplify psychological unreality, people do indeed perform such socially normative behavior: they stop at red lights, pay for their groceries, and remove their hats for the national anthem (Greenwood, 2004). The dilemma with demand characteristics is only that the researcher typically intends for the findings to address a phenomenon that is not simply obedience or expectancy confirmation. The core task for scholars is to properly define the types of situations and settings to which a set of psychological processes generalize, and the theories identified above can aid researchers in precisely this task.

### *Future directions*

The studies reviewed in this manuscript suggest that – specifically within the domain of initial attraction – the hypothetical versus live distinction covaries with forecasted versus experienced emotions, high-level versus low-level construals, and deliberative versus implemental mindsets. What does this observation imply with respect to studies that have used hypothetical paradigms to test hypotheses in domains unrelated to attraction? Sometimes, scholars’ choice of paradigm is likely to make a difference: Although people tend to attribute positive events to their own efforts and negative events to other causes, one meta-analysis revealed that this self-serving bias

is stronger when the events are hypothetical than when they are real (Mezulis, Abramson, Hyde, & Hankin, 2004). The mechanism underlying this difference could indeed be related to the three psychological processes described above, although these mechanisms have not yet been explored to our knowledge.

Yet in other cases, hypothetical paradigms may produce results that are identical to real-life ones: Women are more likely to exhibit forgiveness than men, and this meta-analytic sex difference is just as strong for hypothetical as it is for actual transgressions (Miller, Worthington, & McDaniel, 2008). Such a finding could indicate that the mental operations that characterize forgiveness typically occupy one end of the affective forecasting, construal level, and mindset dimensions. That is, forgiveness may require that people deliberate about future goals and other broader considerations (Finkel, Rusbult, Kumashiro, & Hannon, 2002; Yovetich & Rusbult, 1994), and therefore people gravitate toward abstraction and deliberation in order to forgive even when the transgression is real. Similarly, if we move beyond the hypothetical versus real stimuli distinction and consider the broader laboratory versus field distinction, many research topics exhibit convergence across paradigms. For example, violence research has demonstrated that individual differences and situational factors tend to predict violence equally strongly in the real world versus the laboratory (Anderson & Bushman, 1997), perhaps because the laboratory studies did an excellent job of reproducing the emotionally evocative, anger-inducing situations common in field studies (i.e., participants were experiencing, not forecasting, in these laboratory studies). In short, the current review generates the hypothesis that when research paradigms differ along one or more of the three dimensions identified in Table 1, they will produce different results, and when paradigms occupy the same approximate location on these dimensions, they will produce similar results.

Finally, we should note that other dimensions beyond the three identified here will likely prove relevant to the hypothetical versus live distinction. For example, people may preferentially use social categories (e.g., stereotypes) to evaluate hypothetical others, whereas they may use piecemeal processing to evaluate others in a face-to-face setting (Fiske, Lin, & Neuberg, 1999; Fiske & Neuberg, 1990). Most of the studies in this tradition happened to use hypothetical paradigms and demonstrated that participants defaulted to categorical over piecemeal processing (e.g., Fiske, Neuberg, Beattie, & Milberg, 1987), but perhaps the dominance of categorical processing would weaken or even reverse in the context of a live interaction. In fact, this hypothesis is consistent with the classic LaPiere (1934) study in which restaurant and hotel managers exhibited racist attitudes when asked whether they would serve a hypothetical Asian couple, yet they typically served without prejudice an actual Asian couple who visited their establishments. Similarly, people respond differently to general versus specific cases (Sherman, Beike, & Ryalls, 1999), and it is possible that a live interaction emphasizes the qualities of a specific target, whereas a hypothetical other highlights the qualities shared by a general category of targets. Finally, decision making research has documented that hypothetical versus actual decisions can differ for a variety of reasons, some of which may be distinct from the theories reviewed here (e.g., comprehension of the task; Kühberger, Schulte-Mecklenbeck, & Perner, 2002). No studies have drawn from these frameworks to examine differences between hypothetical and live initial interactions, but they are promising guides for additional research.

## Conclusion

Allport (1968) defined social psychology as the study of “how the thought, feeling, and behavior of individuals are influenced by the actual, imagined, or implied presence of others” (p. 3). Today, the study of actual others and the study of imagined/implied others still belong

under the same epistemological roof, but that does not mean that these different others are psychologically equivalent, interchangeable parts. Some psychological processes may not generalize between evaluations of live versus hypothetical others, and the possibility that people exhibit processing differences across laboratory and real-world contexts has always been the critical question underlying vague concerns about our field's external validity (Mook, 1983). Recent research on attraction has advanced this discussion by positioning such differences within strong theoretical frameworks such as affective forecasting, construal level theory, and mindset theory. The dimensions that underlie these theories are "valid" in the sense that they characterize real psychological phenomena, but participants may be more likely to occupy a particular portion of these dimensions depending on whether they are currently in the laboratory or going about their daily lives. This is a new, more precise way of thinking about the age-old question of external validity – a perspective that may help mitigate the midnight angst of future scholars who seek peace between the rival factions of field and laboratory researchers.

### ***Acknowledgment***

We thank Eli Finkel for his feedback on a previous version of this manuscript. This research was facilitated by National Science Foundation Grant # 1147828 awarded to Paul W. Eastwick.

### **Short Biographies**

Paul Eastwick is an Assistant Professor at the University of Texas at Austin in the Department of Human Development and Family Sciences. His research examines initial attraction, relationship maintenance, and the evolutionary psychology of human mating. He recently was awarded a National Science Foundation grant to study how attachment bonds between mating partners intersect with hormonal influences on mating behavior. His research on the role of ideal partner preferences in predicting people's desire to initiate and maintain romantic relationships received the Society for Experimental Social Psychology (SESP) Dissertation Award.

Lucy Hunt is a graduate student in the Department of Human Development and Family Sciences at the University of Texas at Austin. Her research investigates the extent to which perceivers achieve consensus about the desirable qualities of potential romantic partners. She is also interested in applying these insights to the online dating domain.

Lisa Neff is an Assistant Professor at the University of Texas at Austin in the Department of Human Development and Family Sciences. Her expertise and theoretical interests lie in the study of relationship quality and development during the early years of marriage. She is currently funded by a National Science Foundation grant examining how and when stressors external to the relationship (e.g., work stress, financial difficulties) hinder couples' efforts to engage in relationship-promoting behaviors. She received the Caryl E. Rusbult Early Career Award from the Relationship Researchers Interest Group within the Society for Personality and Social Psychology in 2012.

### **Endnotes**

\* Correspondence address: Department of Human Development and Family Sciences, University of Texas at Austin, Austin, TX 78712, USA. Email: eastwick@austin.utexas.edu

<sup>1</sup> The studies reviewed in this article specifically examined the hypothetical versus live interaction distinction, which has historically been considered a subtype of the broader laboratory versus field distinction. In fact, both Anderson et al. (1999) and Mitchell (2012) conceptualized meta-analyses comparing hypothetical versus real stimuli as examples of

the laboratory versus field comparison. Of course, this analogy is far from perfect; people sometimes encounter descriptions of hypothetical others outside of the laboratory (e.g., online dating profiles). Furthermore, as the studies reviewed below demonstrate, laboratory designs can reproduce the live face-to-face interactions that permeate the real world.

<sup>2</sup> Strictly speaking, implemental and deliberative mindsets are conceptually independent goal pursuit strategies. However, the two mindsets typically do not coexist at a given moment in time but rather function in a sequence like a stage model (Gollwitzer, 2012; Gollwitzer & Bayer, 1999). Thus, for the purposes of the current discussion, implemental and deliberative mindsets can be conceptualized as occupying opposite ends of a “mindset” dimension.

<sup>3</sup> For the purposes of the current discussion, the critical distinction is whether the participants’ experience with the target is an abstract description or a concrete live experience. Although an online dating partner may literally be real, from the perspective of the participant, he/she exists as a hypothetical “profile” until some form of interpersonal interaction takes place.

## References

- Allport, G. (1968). The historical background of modern social psychology. In G. Lindzey & E. Aronson (Eds.), *Handbook of Social Psychology* (2nd ed., Vol. 1, pp. 1–80). Reading, MA: Addison-Wesley.
- Anderson, C. A., & Bushman, B. J. (1997). External validity of “trivial” experiments: The case of laboratory aggression. *Review of General Psychology*, *1*, 19–41.
- Anderson, C. A., Lindsay, J. J., & Bushman, B. J. (1999). Research in the psychological laboratory: Truth or triviality? *Current Directions in Psychological Science*, *8*, 3–9.
- Aronson, E., Wilson, T. D., & Akert, R. (1994). *Social Psychology: The Heart and the Mind*. New York, NY: Harper Collins.
- Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, *2*, 396–403.
- Brewer, M. B. (2000). Research design and issues of validity. In H. T. Reis & C. M. Judd (Eds.), *Handbook of Research Methods in Social and Personality Psychology* (pp. 3–16). Cambridge, UK: Cambridge University Press.
- Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, *54*, 297–312.
- Campbell, L., Simpson, J., Kashy, D., & Fletcher, G. (2001). Ideal standards, the self, and flexibility of ideals in close relationships. *Personality and Social Psychology Bulletin*, *27*, 447–462.
- Carlson, R. (1984). What’s social about social psychology? Where’s the person in personality research? *Journal of Personality and Social Psychology*, *47*, 1304–1309.
- Cialdini, R. B. (2009). We have to break up. *Perspectives on Psychological Science*, *4*, 5–6.
- Dhami, M. K., Hertwig, R., & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychological Bulletin*, *130*, 959–988.
- Eagly, A. H., Eastwick, P. W., & Johannesen-Schmidt, M. (2009). Possible selves in marital roles: The impact of the anticipated division of labor on the mate preferences of women and men. *Personality and Social Psychology Bulletin*, *35*, 403–414.
- Eagly, A. H., & Johnson, B. T. (1990). Gender and leadership style: A meta-analysis. *Psychological Bulletin*, *108*, 233–256.
- Eastwick, P. W. (2009). Beyond the Pleistocene: Using phylogeny and constraint to inform the evolutionary psychology of human mating. *Psychological Bulletin*, *135*, 794–821.
- Eastwick, P. W., Eagly, A. H., Finkel, E. J., & Johnson, S. E. (2011). Implicit and explicit preferences for physical attractiveness in a romantic partner: A double dissociation in predictive validity. *Journal of Personality and Social Psychology*, *101*, 993–1011.
- Eastwick, P. W., & Finkel, E. J. (2008). Sex differences in mate preferences revisited: Do people know what they initially desire in a romantic partner? *Journal of Personality and Social Psychology*, *94*, 245–264.
- Eastwick, P. W., Finkel, E. J., & Eagly, A. H. (2011). When and why do ideal partner preferences affect the process of initiating and maintaining romantic relationships? *Journal of Personality and Social Psychology*, *101*, 1012–1032.
- Eastwick, P. W., Luchies, L. B., Finkel, E. J., & Hunt, L. L. in press. The predictive validity of ideal partner preferences: A review and meta-analysis. *Psychological Bulletin*.
- Festinger, L., Schachter, S., & Back, K. W. (1950). *Social Pressures in Informal Groups: A Study of Human Factors in Housing*. New York: Harper & Brothers.
- Finkel, E. J., Eastwick, P. W., Karney, B. R., Reis, H. T., & Sprecher, S. (2012). Online dating: A critical analysis from the perspective of psychological science. *Psychological Science in the Public Interest*, *13*, 3–66.
- Finkel, E. J., Rusbult, C. E., Kumashiro, M., & Hannon, P. A. (2002). Dealing with betrayal in close relationships: Does commitment promote forgiveness? *Journal of Personality and Social Psychology*, *82*, 956–974.
- Fiske, S. T., Lin, M., & Neuberg, S. (1999). The continuum model: Ten years later. In S. Chaiken & Y. Trope (Eds.), *Dual-Process Theories in Social Psychology* (pp. 231–254). New York, NY: Guilford Press.

- Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 23, pp. 1–74). New York, NY: Academic Press.
- Fiske, S. T., Neuberg, S. L., Beattie, A. E., & Milberg, S. J. (1987). Category-based and attribute-based reactions to others: Some informational conditions of stereotyping and individuating processes. *Journal of Experimental Social Psychology*, 23, 399–427.
- Frost, J. H., Chance, Z., Norton, M. I., & Ariely, D. (2008). People are experience goods: Improving online dating with virtual dates. *Journal of Interactive Marketing*, 22, 51–61.
- Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, 102, 652–670.
- Gagné, F. M., & Lydon, J. E. (2001). Mind-set and close relationships: When bias leads to (in)accurate predictions. *Journal of Personality and Social Psychology*, 81, 85–96.
- Gagné, F. M., & Lydon, J. E. (2004). Bias and accuracy in close relationships: An integrative review. *Personality and Social Psychology Review*, 8, 322–338.
- Gagné, F. M., Lydon, J. E., & Bartz, J. A. (2003). Effects of mindset on the predictive validity of relationship constructs. *Canadian Journal of Behavioural Science*, 35, 292–304.
- Gilbert, D. T., Pinel, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. P. (1998). Immune neglect: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology*, 75, 617–638.
- Gollwitzer, P. M. (1990). Action phases and mind-sets. In E. T. Higgins & R. M. Sorrentino (Eds.), *Handbook of Motivation and Cognition: Foundations of Social Behavior* (Vol. 2, pp. 53–92). New York, NY: Guilford Press.
- Gollwitzer, P. M. (2012). Mindset theory of action phases. In P. A. M. Van Lange, A. W. Kruglanski & E. T. Higgins (Eds.), *Handbook of Theories of Social Psychology* (pp. 526–545). London: Sage Publications.
- Gollwitzer, P. M., & Bayer, U. (1999). Deliberative versus implemental mindsets in the control of action. In S. Chaiken & Y. Trope (Eds.), *Dual-Process Theories in Social Psychology* (pp. 403–422). New York, NY: Guilford Press.
- Greenwood, J. D. (2004). What happened to the “social” in social psychology? *Journal for the Theory of Social Behaviour*, 34, 19–34.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world. *The Behavioral and Brain Sciences*, 33, 61–83.
- Henry, P. J. (2008). College sophomores in the laboratory redux: Influences of a narrow data base on social psychology’s view of the nature of prejudice. *Psychological Inquiry*, 19, 49–71.
- Kawakami, K., Dunn, E., Karmali, F., & Dovidio, J. F. (2009). Mispredicting affective and behavioral responses to racism. *Science*, 323, 276–278.
- Kenrick, D. T. (1986). How strong is the case against contemporary social and personality psychology? A response to Carlson. *Journal of Personality and Social Psychology*, 50, 839–844.
- Kühberger, A., Schulte-Mecklenbeck, M., & Perner, J. (2002). Framing decisions: Hypothetical and real. *Organizational Behavior and Human Decision Processes*, 89, 1162–1175.
- LaPiere, R. T. (1934). Attitudes vs. actions. *Social Forces*, 13, 230–237.
- Ledgerwood, A., Trope, Y., & Chaiken, S. (2010). Flexibility now, consistency later: Psychological distance and construal shape evaluative responding. *Journal of Personality and Social Psychology*, 99, 32–51.
- Ledgerwood, A., Trope, Y., & Liberman, N. (2010). Flexibility and consistency in evaluative responding: The function of construal level. In M. P. Zanna & J. M. Olson (Eds.), *Advances in Experimental Social Psychology* (Vol. 43, pp. 257–295). New York, NY: Elsevier.
- Liberman, N., Trope, Y., & Stephan, E. (2007). Psychological distance. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social Psychology: Handbook of Basic Principles* (2nd ed., pp. 353–381). New York, NY: Guilford Press.
- Loewenstein, G. (1996). Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes*, 65, 272–292.
- Loewenstein, G. (2005). Hot-cold empathy gaps and medical decision making. *Health Psychology*, 24, S49–S56.
- MacDonald, T. K., & Ross, M. (1999). Assessing the accuracy of predictions about dating relationships: How and why do lovers’ predictions differ from those made by observers? *Personality and Social Psychology Bulletin*, 25, 1417–1429.
- Mallett, R. K., Wilson, T. D., & Gilbert, D. T. (2008). Expect the unexpected: Failure to anticipate similarities leads to an intergroup forecasting error. *Journal of Personality and Social Psychology*, 94, 265–277.
- Maniaci, M. R., Reis, H. T., Caprariello, P. A., Eastwick, P. W., & Finkel, E. J. (2012). Reconsidering the role of familiarity in interpersonal attraction. Paper presented at the Interactional Association for Relationship Research Biannual Conference, Chicago, IL.
- McFarland, C., Ross, M., & DeCourville, N. (1989). Women’s Theories of menstruation and biases in recall of menstrual symptoms. *Journal of Personality and Social Psychology*, 57, 522–531.
- Mezulis, A. H., Abramson, L. Y., Hyde, J. S., & Hankin, B. L. (2004). Is there a universal positivity bias in attributions? A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias. *Psychological Bulletin*, 130, 711–747.



- Miller, A. J., Worthington, E. L., & McDaniel, M. A. (2008). Gender and forgiveness: A meta-analytic review and research agenda. *Journal of Social and Clinical Psychology, 27*, 843–876.
- Mitchell, G. (2012). Revisiting truth or triviality: The external validity of research in the psychological laboratory. *Perspectives on Psychological Science, 7*, 109–117.
- Mithen, S. J. (1996). *The Prehistory of the Mind: The Cognitive Origins of art, Religion, and Science*. London: Thames and Hudson.
- Mook, D. G. (1983). In defense of external invalidity. *American Psychologist, 38*, 379–387.
- Moreland, R. L., & Beach, S. R. (1992). Exposure effects in the classroom: The development of affinity among students. *Journal of Experimental Social Psychology, 28*, 255–276.
- Murray, S. L., Holmes, J. G., & Griffin, D. W. (1996). The self-fulfilling nature of positive illusions in romantic relationships: Love is not blind, but prescient. *Journal of Personality and Social Psychology, 71*, 1155–1180.
- Norton, M. I., Frost, J. H., & Ariely, D. (2007). Less is more: The lure of ambiguity, or why familiarity breeds contempt. *Journal of Personality and Social Psychology, 92*, 97–105.
- Nussbaum, S., Trope, Y., & Liberman, N. (2003). Creeping dispositionism: The temporal dynamics of behavior prediction. *Journal of Personality and Social Psychology, 84*, 485–497.
- Park, L. E., Young, A. F., & Eastwick, P. W. (2013). (Psychological) distance makes the heart grow fonder: Effects of psychological distance and relative intelligence on men's attraction to women. *Manuscript Under Review*.
- Pleban, R., & Tesser, A. (1981). The effects of relevance and quality of another's performance on interpersonal closeness. *Social Psychology Quarterly, 278–285*.
- Reis, H. T., & Gosling, S. D. (2010). Social psychological methods outside the laboratory. In S. T. Fiske, D. T. Gilbert & G. Lindzey (Eds.), *Handbook of Social Psychology* (5th ed., Vol. 1, pp. 82–114). Hoboken, NJ: John Wiley & Sons.
- Reis, H. T., Maniaci, M. R., Caprariello, P. A., Eastwick, P. W., & Finkel, E. J. (2011). Familiarity does indeed promote attraction in live interaction. *Journal of Personality and Social Psychology, 101*, 557–570.
- Rim, S. Y., Uleman, J. S., & Trope, Y. (2009). Spontaneous trait inference and construal level theory: Psychological distance increases nonconscious trait thinking. *Journal of Experimental Social Psychology, 45*, 1088–1097.
- Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin, 128*, 934–960.
- Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology, 51*, 515–530.
- Shelton, J. N. (2000). A reconceptualization of how we study issues of racial prejudice. *Personality and Social Psychology Review, 4*, 374–390.
- Sherman, S. J., Beike, D. R., & Ryalls, K. R. (1999). Dual-processing accounts of inconsistencies in responses to general versus specific cases. In S. Chaiken & Y. Trope (Eds.), *Dual-Process Theories in Social Psychology* (pp. 203–230). New York, NY: Guilford Press.
- Simpson, J. A., Campbell, B., & Berscheid, E. (1986). The association between romantic love and marriage: Kephart (1967) twice revisited. *Personality and Social Psychology Bulletin, 12*, 363–372.
- Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103*, 193–210.
- Tesser, A. (1988). Toward a self-evaluation maintenance model of social behavior. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 21, pp. 181–227). San Diego, CA: Academic Press.
- Trope, Y., & Liberman, N. (2003). Temporal construal. *Psychological Review, 110*, 403–420.
- Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review, 117*, 440–463.
- Wilson, T. D., & Gilbert, D. T. (2003). Affective forecasting. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (pp. 345–411). New York: Elsevier.
- Wilson, T. D., Laser, P. S., & Stone, J. I. (1982). Judging the predictors of one's own mood: Accuracy and the use of shared theories. *Journal of Experimental Social Psychology, 18*, 537–556.
- Woodzicka, J. A., & LaFrance, M. (2001). Real versus imagined gender harassment. *Journal of Social Issues, 57*, 15–30.
- Yovetich, N. A., & Rusbult, C. E. (1994). Accommodative behavior in close relationships: Exploring transformation of motivation. *Journal of Experimental Social Psychology, 30*, 138–164.